Feb 5th 2019 Vincent Mai

Uncertainty RL – Experiments on the multi-arm bandits setup

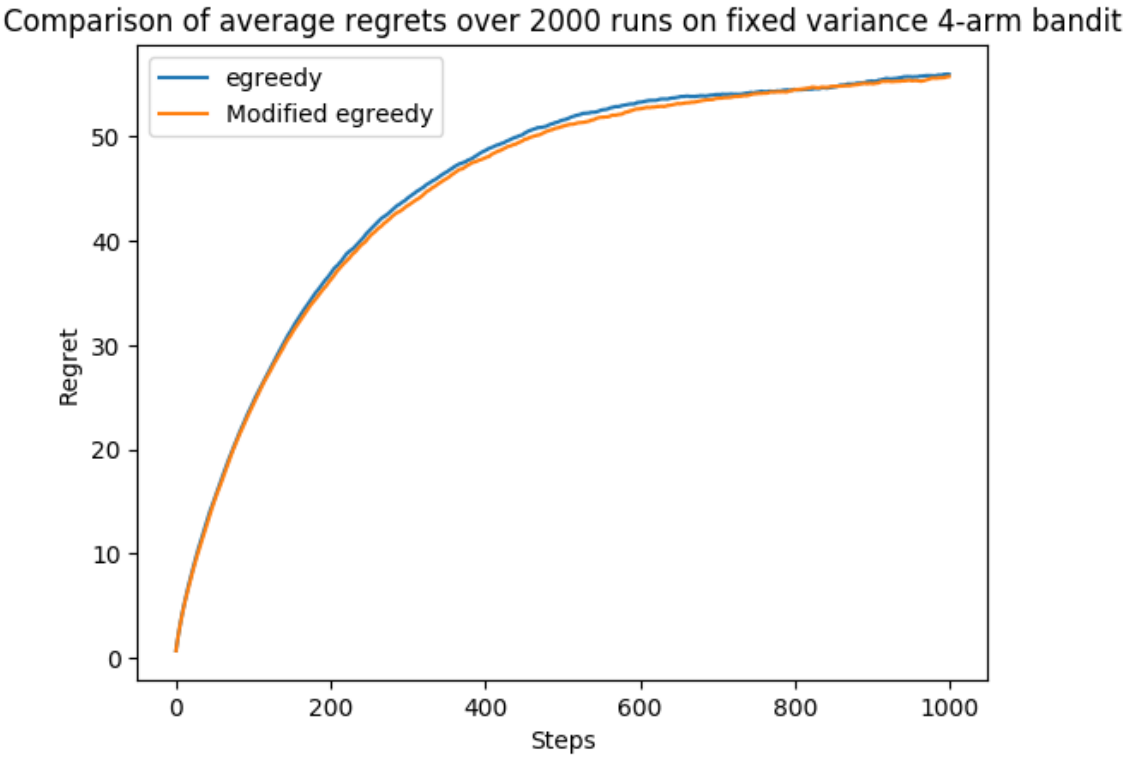
Phase 1 : Run both algos with fixed variance v

1.a.
Environment : v = 0.6, means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algos : e-greedy, modified e-greedy with weights 0.1 + 0.54/var (= 1 here)

Results :



1.b

Environment : v = 0.3, means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algos : e-greedy, modified e-greedy with weights 0.1 + 0.54/var (= 1.8 here)

Results :

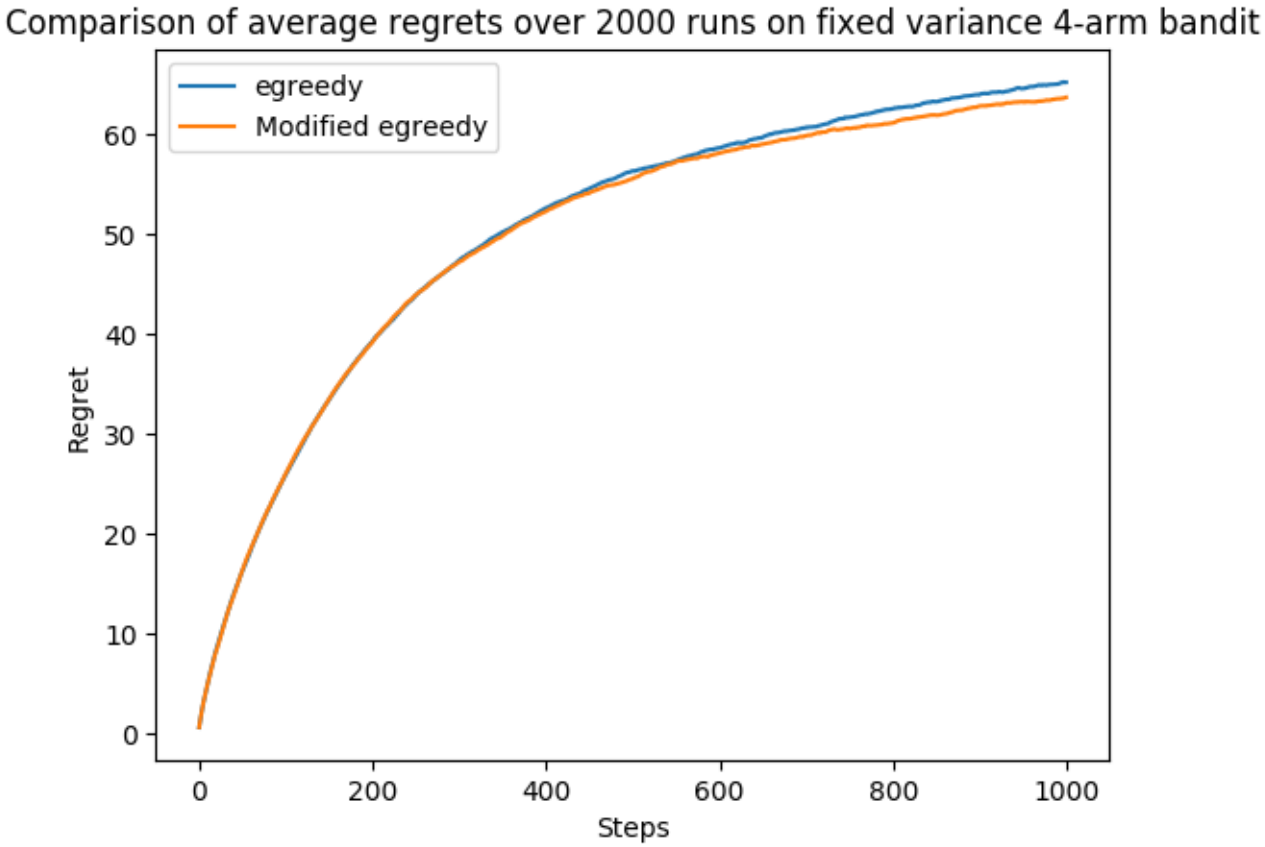
1.c

Environment : v = 1, means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algos : e-greedy, modified e-greedy with weights 0.1 + 0.54/var (= 1.8 here)

Results :

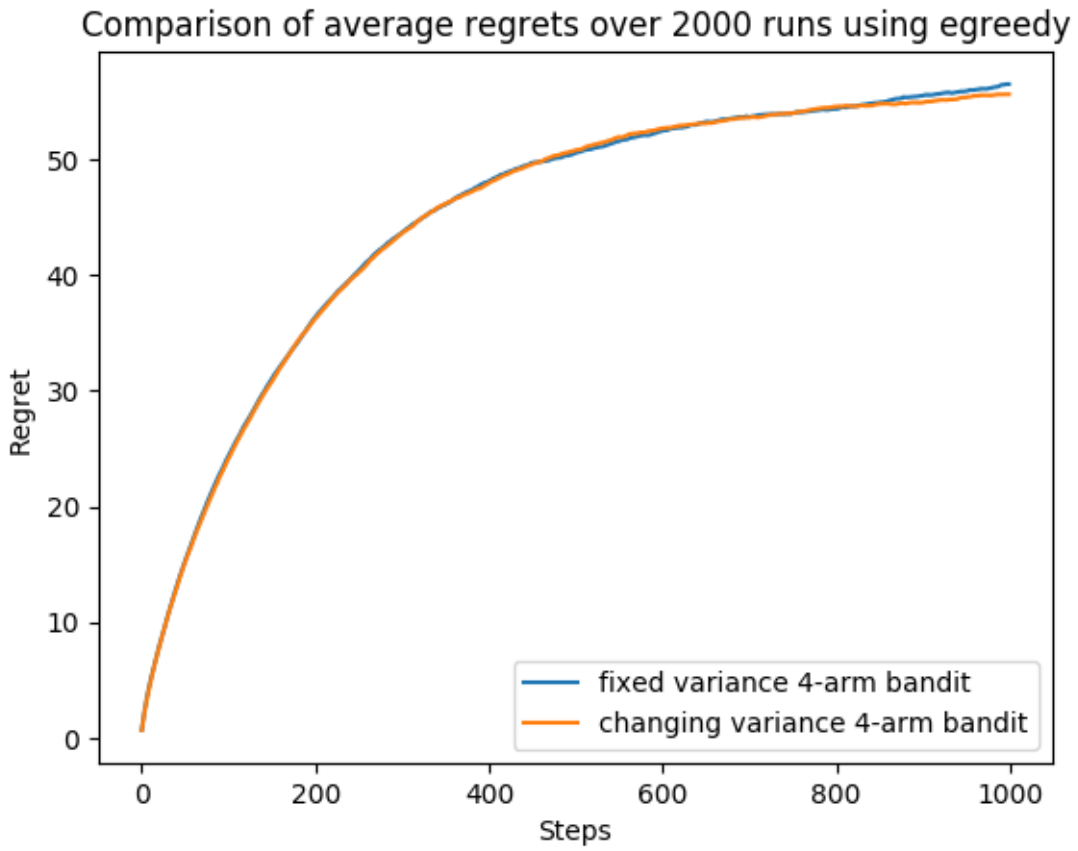


Phase 2 : Compare behavior of one algo in both environments

2.a
Environment 1 : v = 0.6, means = [0, 0.3, 0.6, 0.9]
Environment 2 : v ~ ChiSquare(3) /6 + 0.1 (expected value : 0.6), means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algo: e-greedy



2.b

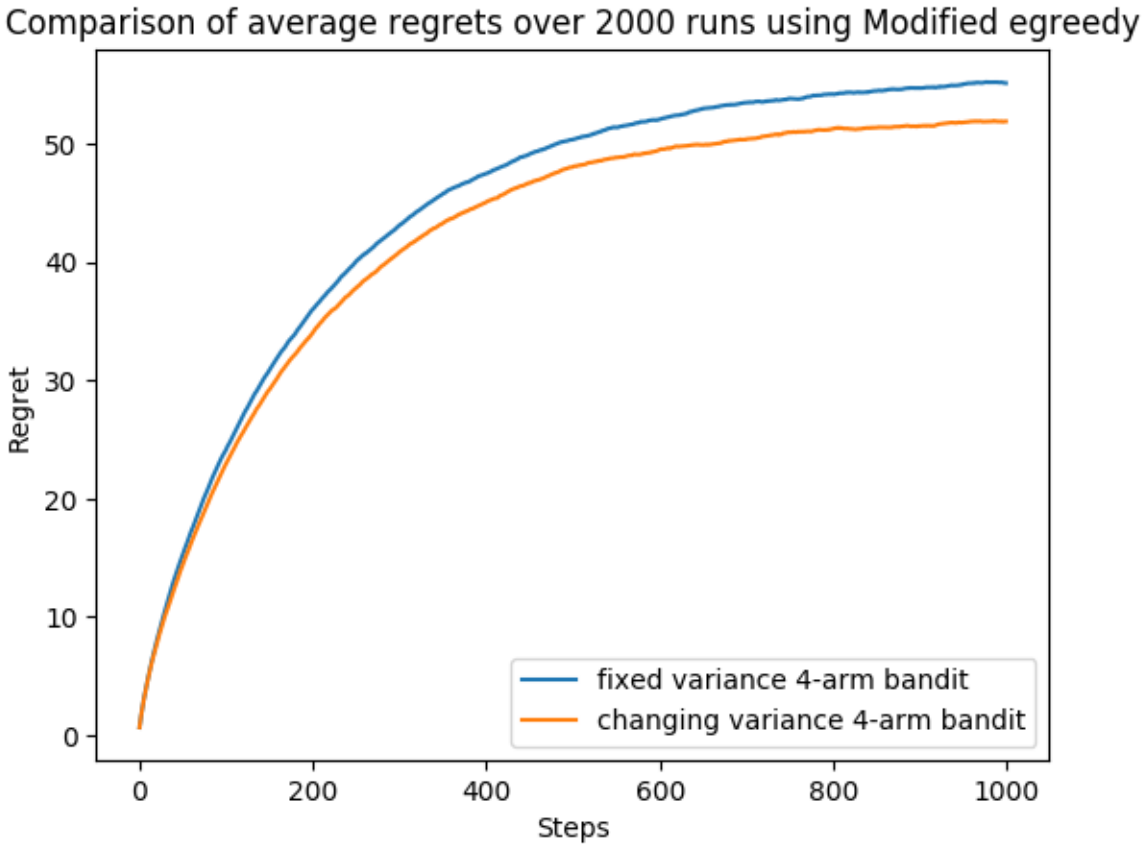
Environment 1 : v = 0.6, means = [0, 0.3, 0.6, 0.9]

Environment 2 : v ~ ChiSquare(3) /6 + 0.1 (expected value : 0.6), means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algo: modified e-greedy with weights 0.1 + 0.54/var (= 1 here)

Results :



Phase 3 : Compare behavior of both algos in changing variance environments

3.a

Environment : v ~ ChiSquare(3) /6 + 0.1 (expected value : 0.6), means = [0, 0.3, 0.6, 0.9]

Steps : 1000 Runs : 2000 Epsilon$_0$: 0.5 Decay : 0.995 Epsilon_min : 0.01

Algos : e-greedy, modified e-greedy with weights 0.1 + 0.54/var (= 1 here)